



Keeney, E., Thom, H. H. Z., Turner, E. L., Martin, R. M., & Sanghera, S. (2021). Using a Modified Delphi Approach to Gain Consensus on Relevant Comparators in a Cost-Effectiveness Model: Application to prostate cancer screening. *PharmacoEconomics*, 39(5), 589-600.
<https://doi.org/10.1007/s40273-021-01009-6>

Publisher's PDF, also known as Version of record

License (if available):
CC BY

Link to published version (if available):
[10.1007/s40273-021-01009-6](https://doi.org/10.1007/s40273-021-01009-6)

[Link to publication record in Explore Bristol Research](#)
PDF-document

This is the final published version of the article (version of record). It first appeared online via Springer Nature at <https://doi.org/10.1007/s40273-021-01009-6>. Please refer to any applicable terms of use of the publisher.

University of Bristol - Explore Bristol Research

General rights

This document is made available in accordance with publisher policies. Please cite only the published version using the reference above. Full terms of use are available:
<http://www.bristol.ac.uk/red/research-policy/pure/user-guides/ebr-terms/>



Using a Modified Delphi Approach to Gain Consensus on Relevant Comparators in a Cost-Effectiveness Model: Application to Prostate Cancer Screening

Edna Keeney¹ · Howard Thom^{1,2} · Emma Turner³ · Richard M. Martin^{2,3,4} · Sabina Sanghera¹

Accepted: 15 February 2021
© The Author(s) 2021

Abstract

Objective Challenges can exist when framing the decision question in a cost-effectiveness analysis, particularly when there is disagreement among experts on relevant comparators. Using prostate cancer screening and recent developments in risk stratification, early-detection biomarkers, and diagnostic technologies as a case study, we report a modified Delphi approach to handle decision-question uncertainty.

Methods The study involved two rounds of anonymous online questionnaires to identify the prostate cancer screening strategies that international researchers, clinicians and decision makers felt important to consider in a cost-effectiveness model. Both purposive and snowball sampling were used to recruit experts. The questionnaire was based on a review of the literature and was piloted for language, comprehension and ease of use prior to dissemination. In Round 1, respondents indicated their preferred screening strategy (including no screening) through a series of multiple-choice questions. The responses informed a set of 13 consensus statements, which respondents ranked their agreement with on a 9-point Likert scale (Round 2). Consensus was considered reached if > 70% of participants indicated agreement and < 15% indicated disagreement.

Results Twenty participants completed Round 1 and 17 completed Round 2. Consensus was shown towards comparing no formal screening, age-based, and risk-based strategies. The risk-based approaches included screening only higher-risk men, using shorter screening intervals for higher-risk men, screening higher-risk men at an earlier age, and tailoring screening intervals based on prostate-specific antigen (PSA) level at a previous test. There was agreement that inclusion of MRI in the pathway should be considered, but disagreement on the inclusion of new biomarkers.

Conclusion In disease areas where technologies are rapidly evolving, a modified Delphi approach provides a useful tool to identify relevant comparators in an economic evaluation.

1 Introduction/Background

To determine whether new healthcare innovations should be implemented at a national level, evidence on both long-term effectiveness and cost effectiveness is required.

Generally, it is recommended that cost-effectiveness models include all possible comparators, or clearly justify their exclusion [1–3]. Philips et al. [2] state that this should be “based on literature reviews and expert opinion and will include evidence from local guidelines and treatment patterns”.

As with all health care interventions, cost-effectiveness models for screening interventions include many components that have the potential to change over time. The frequency of screening, population to screen, diagnostic test or tests used, and their order and combination are all subject to changing practice and can make identifying all possible comparators, especially “what is relevant?”, challenging. Prostate cancer screening in particular is an area where considerable uncertainty as to what the comparators in a cost-effectiveness analysis should be exists. There is an ongoing debate around whether to screen and if so, who, how and how often in an era of relatively rapid

✉ Edna Keeney
edna.keeney@bristol.ac.uk

¹ Health Economics Bristol (HEB), Population Health Sciences, Bristol Medical School, University of Bristol, 1-5 Whiteladies Road, Bristol BS8 1NU, UK

² National Institute for Health Research (NIHR) Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol, Bristol, UK

³ Population Health Sciences, Bristol Medical School, University of Bristol, Bristol BS8 2PR, UK

⁴ MRC Integrative Epidemiology Unit, University of Bristol, Bristol, UK

Key Points

This study has demonstrated an application of the modified Delphi approach to determine relevant options for comparison in a cost-effectiveness analysis, where uncertainty exists.

Using prostate cancer as a case study, the method was successful in identifying the prostate cancer screening strategies that are relevant to researchers, clinicians and decision makers.

evolving practice and advancements in screening and testing technologies.

Current practice in the UK, Europe and the US generally consists of opportunistic/unorganised prostate-specific antigen (PSA) testing with a recommendation against formal screening [5–10]. This is because previous methods of relying solely on untargeted PSA testing have not clearly shown that the benefits of early detection using PSA testing outweigh the potential harms of overdiagnosis and overtreatment [11]. However, recent advancements in the identification of risk factors for aggressive disease, such as newly identified genetic variants (polygenic risk) [12]; early detection biomarkers such as the Stockholm 3 panel (STHLM3) [13]; and new strategies to target diagnostic biopsy, including the use of multiparametric magnetic resonance imaging (mpMRI [14]), potentially offer new opportunities to reduce overdiagnosis and overtreatment. The long-term effectiveness and cost effectiveness of these new approaches (individually or in combination) in reducing overdiagnosis and mortality in prostate cancer screening has yet to be ascertained.

The gathering of expert opinion on the most relevant strategies in clinical practice, and therefore the most useful strategies to compare in a cost-effectiveness model, is a recommended first step in many if not all guidelines on conceptualizing decision models [2–4, 15]. However, the methods to do so have not been clearly formulated. Taking current practice of opportunistic screening as the reference case, our aim was to provide a method to handle decision question uncertainty by using a modified Delphi approach to elicit which recent developments in prostate cancer screening are relevant for comparing in a cost-effectiveness analysis and will provide meaningful findings to decision makers.

2 Methods

The Delphi technique provides a systematic method of gaining consensus from a group of experts through collecting and aggregating informed judgments over multiple iterations [16, 17]. Expert participants are asked to provide their opinions in sequential questionnaires, given group feedback from the previous round. Feedback from sequential rounds encourages participants to reassess, alter and/or develop opinions. Responses are anonymous to ensure that no individual dominates the process [18].

Using a modified Delphi design means that, rather than consulting an expert panel to generate the questions addressed in the study, the researcher collects the initial question(s) through some other means, such as a literature review, and presents them to the panel to begin the consensus seeking process [19]. The modified Delphi technique used in our study is described in Fig. 1.

2.1 Participants and Recruitment

In identifying experts, European Food Safety Authority guidance [17] was used, in that the research team first decided what type of expertise was needed to answer the questions (i.e. expertise in prostate cancer early detection, treatment, modelling or the role of genetics in prostate cancer risk), and then attempted to identify individuals with this particular expertise. To identify and recruit experts to take part in the modified Delphi process, purposive sampling was used. This focuses on the views of those able to provide in-depth knowledge of the topic of interest [20]. In-depth knowledge related to factors such as on-the-job experience and peer-reviewed academic output. The participants worked in a variety of institutions including Universities in the UK, Sweden, Finland, the Netherlands and the US, cancer centres in the US, NHS trusts, Public Health England, and the Institute of Cancer Research. The selection of participants from a variety of backgrounds and institutions was important to encourage diversity of opinions. The purposive sampling was complemented with snowball sampling, whereby the contacts were asked to suggest others within or outside their organisations that might offer insight [21]. Prior research experience, clinical focus, country, and age range were included in the questionnaire so that the impact of these factors on preferences could be assessed.

Contacts were first sent a detailed participant information sheet (Supplementary material 1, see electronic supplementary material [ESM]), explaining that the aim of the study was to help the study team understand the prostate cancer screening strategies that are relevant to compare in a

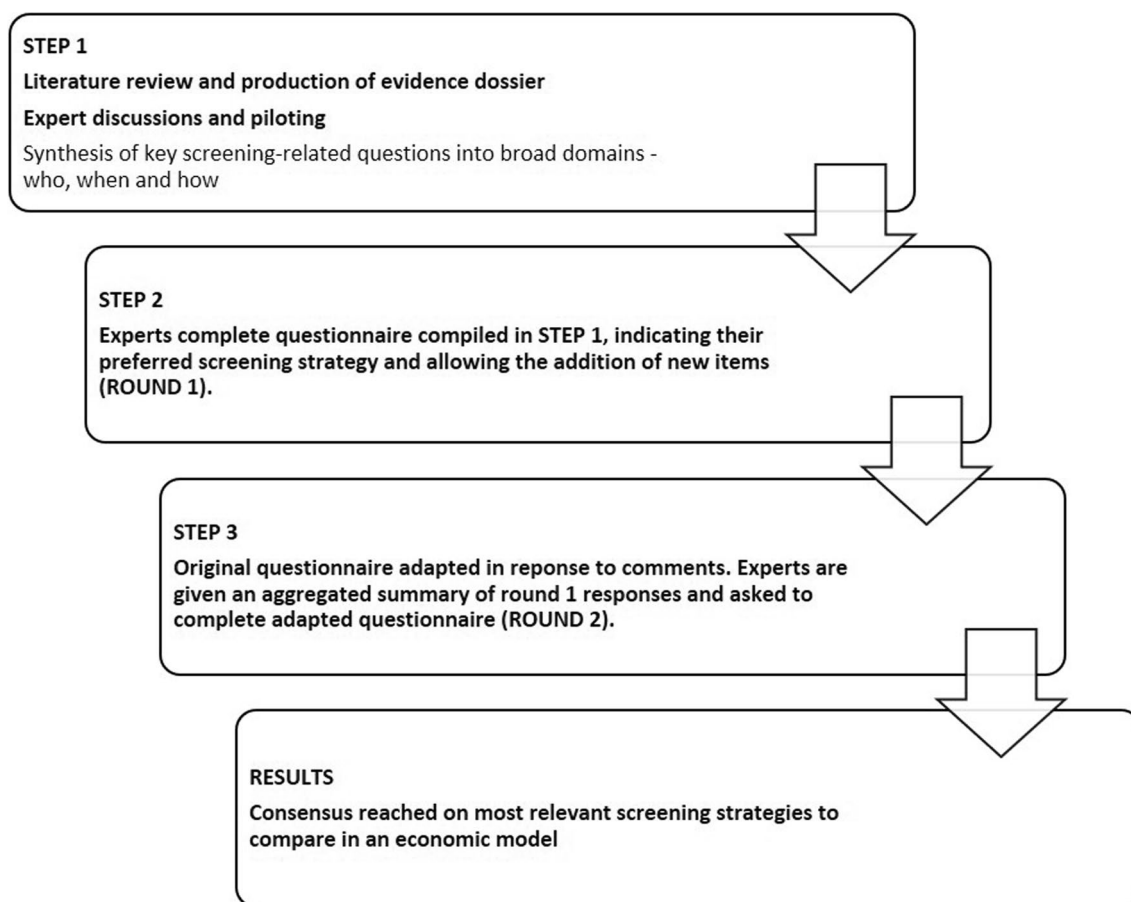


Fig. 1 Modified Delphi technique

cost-effectiveness model, given the rapidly changing landscape in prostate cancer screening. The sheet also informed participants on why they were being asked to take part and what they would be required to do.

2.2 Step 1: Production of Questionnaire and Evidence Dossier

To identify the extent of the uncertainty in the decision question, a rapid review of the literature was carried out. The review involved identifying and summarising current UK and international guidelines on prostate cancer screening in terms of what diagnostic tools to use, the ages to start and stop screening and recommended screening frequency. Recent large trials that addressed different aspects of screening, and which have informed these guidelines, were also identified. An evidence dossier providing an overview of relevant trials was produced using the review evidence (Supplementary Material 2, see ESM). The evidence dossier was circulated to an advisory panel comprising two oncologists, two urologists, one GP, and two researchers in prostate cancer screening and clinical oncology (four from

the UK and three from the US), to identify any missing evidence.

The findings from the review informed the development of the questionnaire, which centred around three primary domains of uncertainty: (1) the group of men that should be invited for screening; (2) how often they should be screened; and (3) which diagnostic procedures to use. These domains were considered basic starting points of any screening strategy with the aim of the questionnaire being to clarify areas of uncertainty within these domains.

Although training was not provided to the participants in advance, the questionnaire was piloted for language, comprehension and ease of use on a smaller group comprising a clinical oncology researcher, a medical oncologist, a GP and a urologist. No difficulty was demonstrated in completing the questionnaire and the answers given indicated a good degree of understanding. Changes were made in response to their comments, such as a move away from a ranking of preferences within each question and towards the selection of one preferred approach. This enabled the experts' preferences among the available options to be more clearly drawn out and was in response

to comments that after a certain rank the experts had no real preference.

2.3 Step 2: Dissemination of Questionnaire

A web-administered version of the questionnaire was developed using REDCap electronic data capture tools hosted at the University of Bristol [22]. Links to the questionnaires and evidence dossier were circulated to the participants via email. The questionnaire was sent to 20 experts in the first instance, with seven additional participants included through snowball sampling. A minimum sample of 23 has been shown to be sufficient and reliable in Delphi surveys when experts are carefully selected [23]. Approval for the study was granted by the Faculty of Health Sciences Research Ethics Committee of the University of Bristol (Approval reference number: 91622).

In Round 1, respondents were asked to indicate their preferred screening strategy through a series of questions. This included an option of ‘no screening’. Completion of these questions resulted in the generation of an automatic statement that summarised each respondent’s answers. An example of this can be seen in Box 1:

“Your preferred screening strategy is age-based, in men aged 50 to 75. You believe the PSA threshold for further investigation should be 3 ng/ml and that the screening interval should be every two years. Men with a raised PSA should be tested using a multi-kallikrein panel (e.g. 4k score, STHLM3) and multiparametric MRI (mpMRI) prior to biopsy.”

Using a free-text section at the end of each question, respondents could add items they considered important but that were not already covered or indicate their uncertainty with regard to a particular question. An example of a completed questionnaire is available in Supplementary Material 3 (see ESM).

2.4 Step 3: Updating of Questionnaire and Evidence Dossier

The first-round responses informed the content and modification of an updated questionnaire (Round 2). At the end of Round 1, the percentage of participants choosing each item was summarised. Items were not considered in Round 2 if not chosen by any participant in Round 1. Nor were they included if chosen by >70% of participants, as consensus was already considered reached. Previous Delphi studies have also used this level of agreement [24]. New items were considered if suggested by > 10% of participants, also in line with a previous Delphi consensus process [25].

An updated questionnaire, in the form of 13 consensus statements, was then sent to the same key contacts from Round 1 and their lists of suggested contacts. The statements were designed to clarify the respondent’s views from Round 1, with an emphasis on the strategies for inclusion in a cost-effectiveness model. Feedback from Round 1 was presented for each Round 2 item in the form of the number of participants choosing each option and a summary of comments given. The first round was completed in November 2019, and the second round in January 2020. As 2 months had passed since the first round, participants were given a reminder of their own choice.

As is common in Delphi procedures [24, 26], respondents were requested to rank their agreement with the 13 statements on a 9-point Likert scale (i.e. ranging from 1 = “strongly disagree” to 9 = “strongly agree”). Consensus on a statement was considered reached if scored 7 to 9 (moderately agree, agree, or strongly agree) by >70% of participants and 1 to 3 (strongly disagree, disagree or moderately disagree) by <15% of all participants [24]. In place of conducting further Delphi rounds, which were not deemed necessary after considering the results of the first two rounds, more stringent criteria were also set (> 70% scoring an item 8 or 9 and <15% scoring 1–3), to aid discussion on the most important aspects to participants, in line with a previous

consensus process [25]. Within-group interrater agreement was assessed using the r^*_{wg} statistic with the rectangular null and maximum dissensus null distributions [27]. This is a commonly used statistic to quantify consensus in ratings on a Likert scale [28]. With this statistic $r^*_{wg} \geq 0.80$ may be considered high enough agreement to establish interrater agreement with 10 or more raters.

3 Results

3.1 Round 1

3.1.1 Characteristics of Respondents

Twenty participants (74%) responded to the questionnaire, out of 27 invitees. The respondents were of varying ages with most being in the 45–64 years age bracket. Nine out of 20 (45%) respondents had over 20 years’ experience in

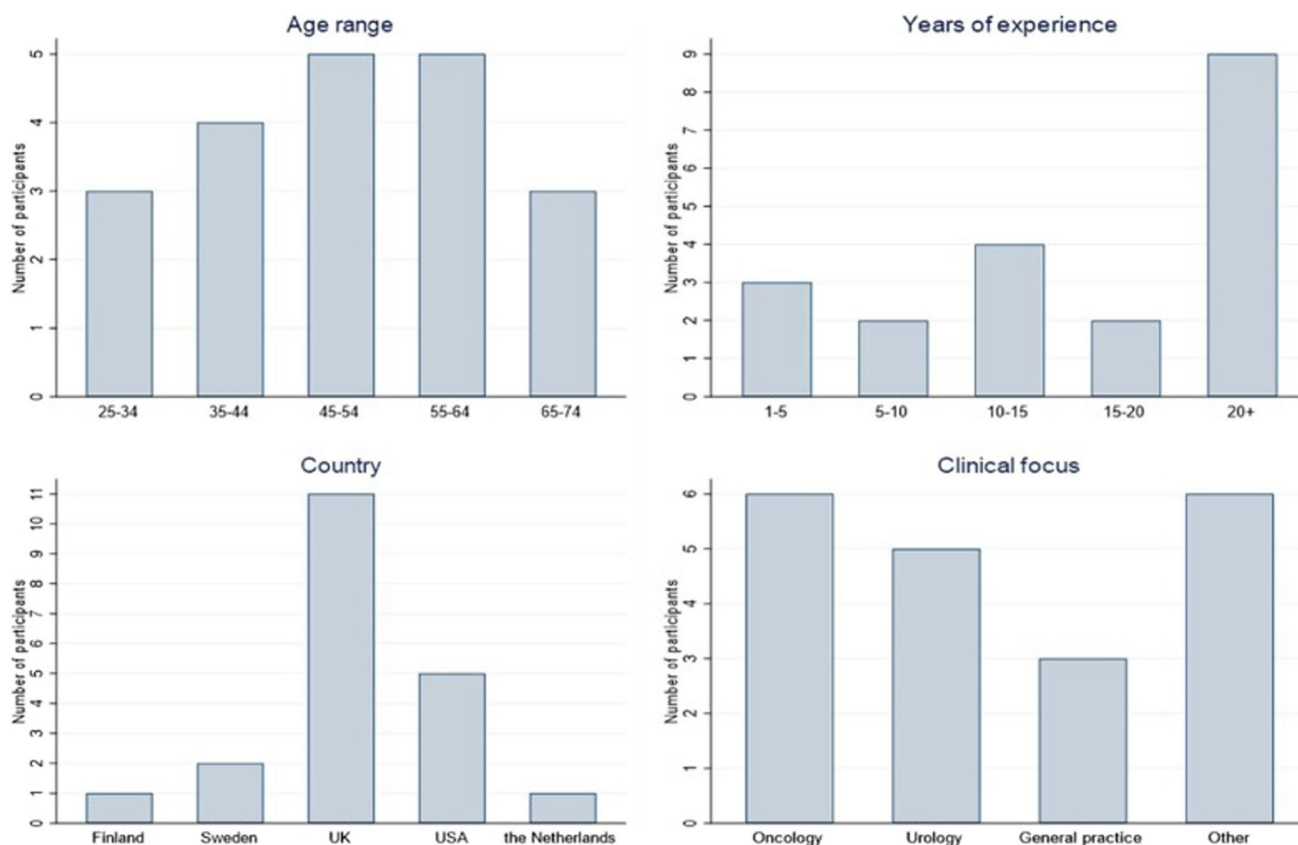


Fig. 2 Characteristics of respondents—Round 1

their field. Responses were received from six oncologists, five urologists, three GPs, two public health specialists, two researchers with experience of modelling prostate cancer, and two people involved in UK national screening services. Four of the respondents had extensive experience in cancer epidemiology, and three of these were Professors in general epidemiology or in cancer epidemiology, specifically. Eleven out of 20 (55%) of the respondents were UK based (Fig. 2). As responses were received from participants with a range of expertise and backgrounds, 20 participants were considered sufficient to collect rich data and allow exploration of the screening strategies thought to be important.

3.1.2 Who Should Be Invited For Screening?

With the aim of determining the most relevant strategies to compare in a cost-effectiveness model, the first question asked what type of screening programme, if any, participants felt should be provided in the UK. Options included no screening, opportunistic screening, organised age-based screening and organised risk-based screening. No participant chose opportunistic screening. The most popular choice was risk-based screening, chosen by 11 participants (55%),

followed by no screening (5/20 participants, 25%) and age-based screening (4/20 participants, 20%) (Fig. 3a).

Risk-based screening was chosen by all five urologists and no screening was chosen by all of those involved in national screening services. No other trends were identified based on age, experience, or country of residence. Reasons given for preferring risk-based screening included “early detection and intervention for people with strong risk factors”, and “more scope for improvement as we learn to better estimate a man’s risk”. Reasons given for selecting age-based screening included preventing opportunistic screening in the “wrong” age ranges, and for simplicity. All participants who selected ‘no screening’ highlighted that this was due to a lack of current evidence that screening offers more benefit than harm.

The second question asked all participants who had indicated risk-based screening to be their preferred option which factor/s they thought risk should be based on. Multiple options could be chosen. Five of 11 participants who chose risk-based screening ticked all available options. The most popular option was family history of prostate cancer (chosen by 9 participants), followed by life expectancy > 10 years (based on a risk calculator) (8),

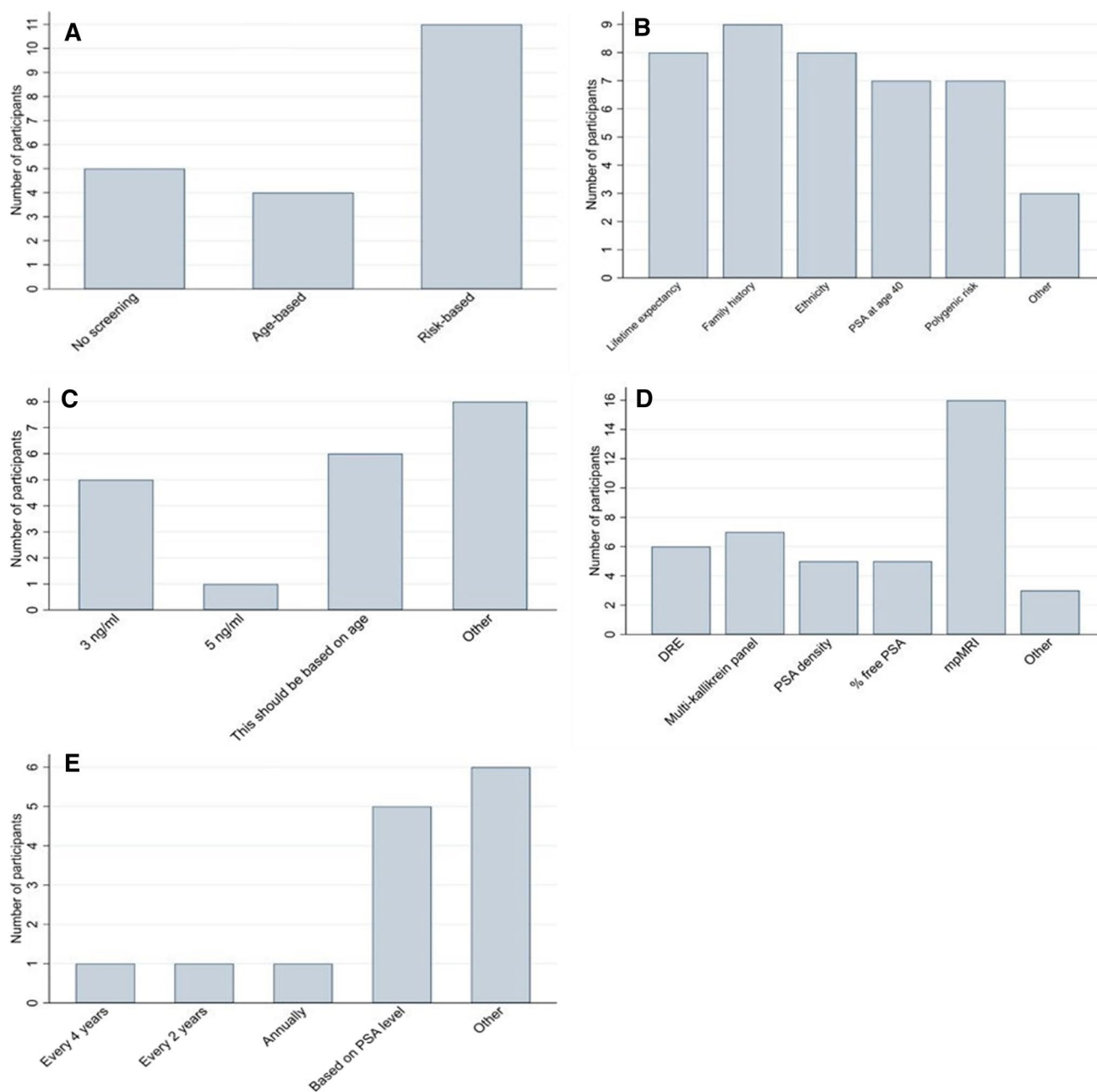


Fig. 3 Number of participants selecting each response to questions, **a** “What type of prostate cancer screening programme do you feel should be provided in the UK?”, **b** “If risk-based screening were to be provided what factor/s should risk be based on?”, **c** “Which PSA threshold do you think should be used to indicate further investiga-

tion?”, **d** “What further investigation should men with a raised PSA level have prior to being offered a biopsy?”, and **e** “Assuming some optimal strategy for inviting men to be screened has been adopted, how frequently do you think men should be screened?”

ethnicity (8), PSA at age 40 years above a certain threshold (7) and polygenic risk (which provides an indication of whether someone might be genetically predisposed to developing prostate cancer) (7), giving an even spread of responses (Fig. 3b). Other suggestions were biomarkers

and “germline high risk mutations”. Comments on this question centred around the understanding that “all available risk factors which can be measured reliably and affordably should be included” with the optimal combination of factors being unclear.

3.1.3 Which Diagnostic Procedures?

The next question asked participants, if a screening programme were to be implemented, which PSA threshold they thought should be used to indicate further investigation, with the options being 3 ng/mL, 4 ng/mL, 5 ng/mL, 6 ng/mL, 7 ng/mL, 8.5 ng/mL, 10 ng/mL, and 'this should be based on age'. Five of 20 participants chose a threshold of 3 ng/mL and one chose a threshold of 5 ng/mL (Fig. 3c). In the 'other' column, two participants suggested that the PSA threshold for investigation should be < 3 ng/mL, two suggested that PSA should be combined with other biomarkers, two suggested that risk calculators should be used in place of PSA and two commented that they did not support any screening.

Comments on this question generally reflected the idea that "a sequence of tests" should be carried out before proceeding to biopsy. Although five people chose a level of 3 ng/mL as their preferred option, three people commented that trials have shown using a level of 3 ng/mL will mean that some aggressive cancers are missed.

The next question asked respondents, if a screening programme were to be implemented, what further investigation(s) they thought men with a raised PSA level should have prior to being offered biopsy. Multiple options could be chosen. The most popular option was mpMRI with 16 respondents (80%) indicating that they thought this should be used prior to biopsy. Seven out of 20 (35%) respondents indicated that a multi-kallikrein panel (e.g. 4kscore, STHLM3) should be used, either with mpMRI alone (4 participants) or alongside PSA density, percent free PSA and mpMRI (3 participants). Six chose digital rectal examination (DRE) as an option (Fig. 3d). Other responses were 'no screening' and 'polygenic markers'. All three GPs indicated that they thought mpMRI should be used prior to biopsy, as did all participants from Sweden, the Netherlands and nine out of ten participants from the UK. All four of those involved in public health or research thought that DRE should be offered.

3.1.4 How Frequently to Screen?

The final question asked the participants who favoured screening over no screening how often they thought men should be screened, given their optimal screening strategy. The options were every 10, 6, 4, and 2 years, annually, only once, and 'this should be based on PSA level'. No participants chose 10 years, 6 years or only once as an appropriate interval and only one participant chose each of 4 years, 2 years and annually. The most popular responses were 'this should be based on PSA level' and 'other' (Fig. 3e).

Of those who thought that screening interval should be based on PSA level, all stated that men with a PSA level <3 ng/mL could be screened every 2–3 years whereas men

with a higher PSA should be screened annually. In the 'other' column, the suggestions made were mainly around screening interval being based on risk.

In summary, the responses to Round 1 indicated several items that were not of interest to the clinical experts surveyed in this sample and therefore assumed to not be of interest to compare in a cost-effectiveness model. These included opportunistic screening, a PSA threshold for further investigation > 3 or a fixed screening interval for all men, as these were not chosen by any participant. On the other hand, a clear consensus (in terms of the pre-defined criteria of >70% agreement) was shown towards the examination of the use of mpMRI prior to biopsy in any potential screening strategy.

3.2 Round 2

The results from Round 1 formed the basis of Round 2, which was a set of 13 consensus statements designed to clarify the expert's responses. As interest was shown in all three of no screening, age-based and risk-based strategies, respondents were again asked their opinion on each but this time in the context of their inclusion as comparators in a cost-effectiveness model. Similarly, further statements attempted to clarify the respondent's views on the inclusion of PSA/other biomarker testing in the screening pathways to be considered in a cost-effectiveness model and how screening intervals should be determined. Seventeen of the 20 participants who completed Round 1 also responded to Round 2, giving an 85% response rate. The participants were again from a wide and representative range of backgrounds (Fig. 4).

The statements that participants were asked to indicate their agreement with are shown in Table 1. This table also shows the percentage of participants rating a statement 7–9 (moderately agree, agree, or strongly agree). A threshold of 70% in this category was the pre-agreed marker of consensus. The percentage of patients rating a statement 8–9 is also shown as well as the percentage rating a statement 1–3 (strongly disagree, disagree or moderately disagree). Detailed results for each statement are given in Supplementary material 4 (see ESM).

Consensus was considered reached, under both the pre-agreed rules (> 70% scoring an item 7–9 and < 15% scoring 1–3) and the more stringent rules (> 70% scoring an item 8 or 9 and < 15% scoring 1–3) on the following five strategies to compare in a cost-effectiveness model:

1. No screening
2. Inviting all men within a certain age range to be screened
3. Inviting only higher risk men for screening (if it is possible to identify higher risk men through the use of polygenic risk scores, family history, ethnicity or otherwise)

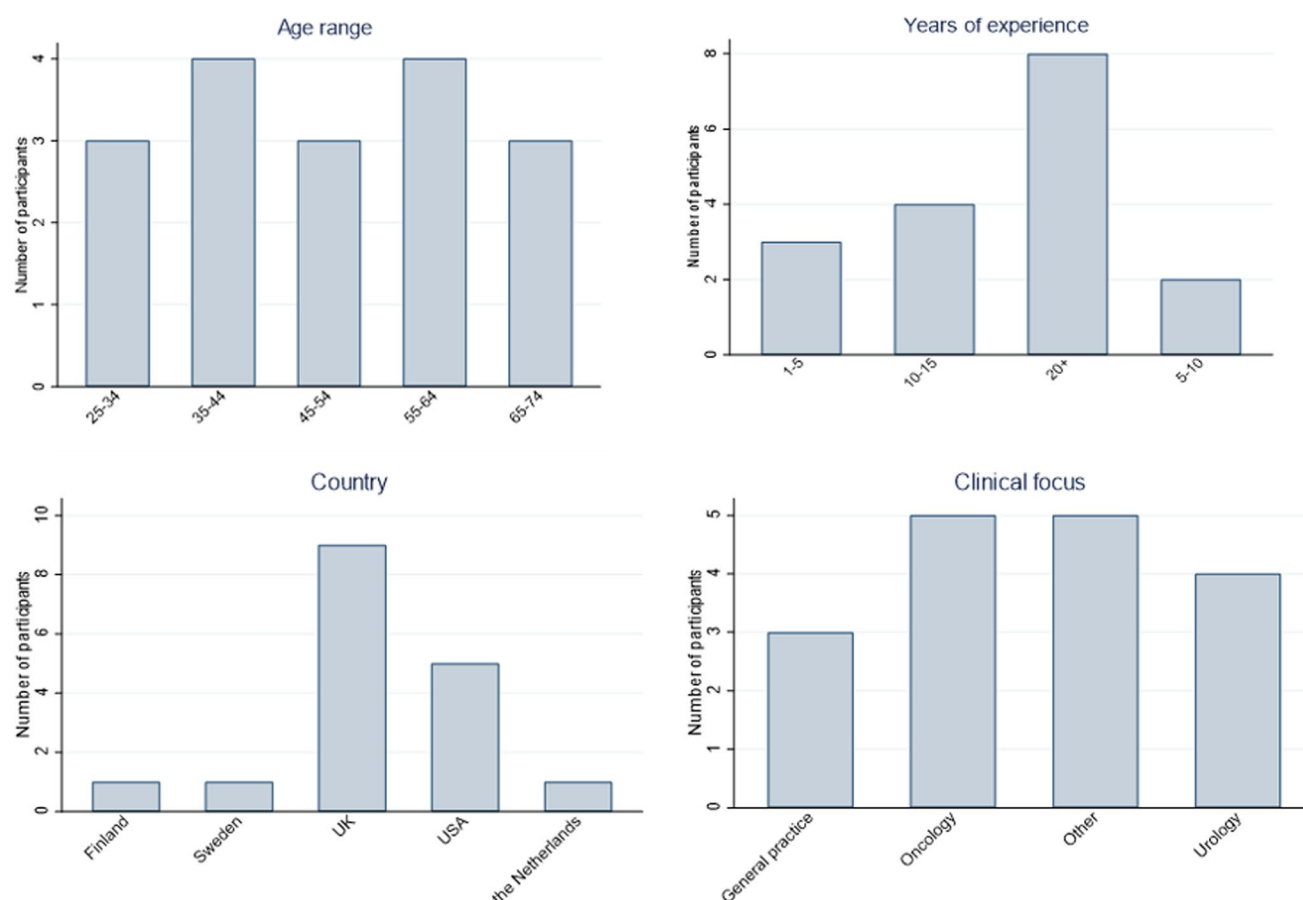


Fig. 4 Characteristics of respondents—Round 2

- Inviting all men within a certain age bracket for screening, but screening higher risk men at an earlier age
- Using different screening intervals for higher and lower risk men

Consensus was also considered reached under the pre-agreed criteria, but not the more stringent criteria, on

- Using different screening intervals based on PSA level at previous test

For all of these strategies, interrater agreement was considered high (>0.8).

Consensus was not considered reached on comparing aspects relating to PSA and other biomarker testing. The responses suggest that participants did not agree on whether PSA in isolation should be used as a reflex test, whether a PSA test should be used alongside a more sophisticated biomarker or risk model, whether a PSA threshold of 1.5 ng/mL has enough negative predictive value to exclude any further testing or whether the threshold for further investigation should increase as men age. In all four of these statements,

over 35% of respondents indicated agreement but at least 5% indicated disagreement. Comments mainly addressed the need for additional and reliable evidence. There was also uncertainty as to the added benefit of biomarkers over the use of MRI.

Although consensus was not considered reached on whether reflex tests such as multi-kallikrein panels, risk models, PSA density or percent free PSA should be evaluated for incorporation in potential screening strategies, there was a tendency towards agreement rather than disagreement with no participants rating either of these statements 1–3. Concerns raised in the comments again centred around a lack of reliable evidence.

With regard to the question of whether all men being screened should be offered a DRE, there was a clear tendency towards disagreement with only 18% of participants being in agreement and 47% disagreeing, suggesting that DRE should not be included when evaluating potential screening strategies. Concerns raised with offering all men a DRE included deterring patients and a high false positive rate.

Table 1 Final outcomes for statements after Round 2

Statement	% Rating 7–9	% Rating 8–9	% Rating 1–3	Outcome: pre-agreed rules	Interrater agreement (r_{wg}^*)
It is useful to compare the cost-effectiveness of no screening to other screening strategies in a future economic model	88.24	88.24	0.00	Consensus reached	0.90
It is useful to compare the cost-effectiveness of inviting all men within a certain age range to be screened to other screening strategies in a future economic model	94.12	82.35	5.88	Consensus reached	0.89
If it is possible to identify men at higher risk of developing prostate cancer prior to testing (through the use of polygenic risk scores, family history, ethnicity or otherwise), it would be useful to compare the cost-effectiveness of inviting only higher risk men for screening	88.24	82.35	0.00	Consensus reached	0.94
If it is possible to identify men at higher risk of developing prostate cancer prior to testing, it would be useful to compare the cost-effectiveness of inviting all men within a certain age bracket for screening but screening higher risk men at an earlier age	88.24	70.59	5.88	Consensus reached	0.82
PSA in isolation should no longer be used a reflex test to trigger MRI/prostate biopsy	35.29	23.53	17.65	Consensus not reached	0.75
A PSA test should be used before a more sophisticated biomarker or risk model (e.g. 4k score, STHLM3) and only men with total PSA above a certain threshold should be tested using the biomarker or risk model	41.18	35.29	23.53	Consensus not reached	0.65
A PSA threshold of 1.5 ng/ml has enough negative predictive value to exclude any further testing	35.29	23.53	17.65	Consensus not reached	0.65
The threshold for further investigation should increase as men age	47.06	35.29	5.88	Consensus not reached	0.76
It would be useful to assess the cost-effectiveness of using a multi-kallikrein panel or risk model (e.g. 4k score, STHLM3) as a reflex test to triage patients suitable for mpMRI prior to biopsy	64.71	52.94	0.00	Consensus not reached	0.87
It would be useful to assess the cost-effectiveness of using PSA density and % free PSA alongside a multi-kallikrein panel as reflex tests to triage patients suitable for mpMRI prior to biopsy	52.94	41.18	0.00	Consensus not reached	0.89
All men being screened should be offered a DRE	17.65	17.65	47.06	Consensus not reached	0.59
If it is possible to identify men at higher risk of developing prostate cancer (through the use of polygenic risk scores, family history, ethnicity or otherwise), it would be useful to compare the cost-effectiveness of using different screening intervals for higher and lower risk men	88.24	82.35	0.00	Consensus reached	0.93
It is useful to compare the cost-effectiveness of using different screening intervals based on PSA level at previous test	76.47	64.71	0.00	Consensus reached*	0.89

DRE digital rectal examination, mpMRI multiparametric magnetic resonance imaging, PSA prostate-specific antigen, STHLM3 Stockholm 3 panel

*This was the only statement for which the outcome under the more stringent rules differed i.e., consensus not reached

4 Discussion

Our aim was to illustrate a method, using prostate cancer screening as an exemplar, to identify strategies to be evaluated in cost-effectiveness modelling when there is considerable uncertainty surrounding the relevant comparators (decision question uncertainty). This process has highlighted the uncertainty and diverging views that can exist and a means to focus these views. We have endeavoured to elicit views from experts working in the field, in accordance with guidelines, and to ensure the findings are applicable to decision makers.

This is the first study to both use and illustrate how to use a modified Delphi method to handle decision question uncertainty and identify relevant comparators for a cost-effectiveness analysis in a rapidly evolving decision-making context. However, the method has previously been used to define the boundaries of a model, in model conceptualisation, and to identify face validity [29–31].

Overall agreement was obtained on the patient characteristics and screening technologies to consider in cost-effectiveness modelling. Although the panel did not reach consensus on exact age ranges to screen or a specific screening interval, future cost-effectiveness models could explore different combinations of screening stopping/starting ages and intervals within the limits suggested by the participants. This could identify the most cost-effective combination of these strategies to enable assessment of whether the ‘optimal combination strategy’ could be cost effective compared with current practice (opportunistic screening).

One strength of the study was the panel of experts, who had a wide range of experience. Relying solely on one project team to identify relevant strategies may have resulted in biased or outdated views. The modified Delphi method provided a systematic way to gain consensus (according to the predefined criteria) from a wide variety of experts. The web-based format enabled the inclusion of views from respondents from geographically dispersed locations. The anonymity meant that no one voice was given precedence and experts had time to consider their responses. An additional advantage was the relative speed of the process, ensuring relevancy in the context of newer innovations.

The aim of any future cost-effectiveness analysis should be to determine, not which screening strategy is most cost effective, but rather if any of the identified screening strategies have the potential to be more cost effective than current practice (i.e. opportunistic PSA-based screening). This can be determined by exploring which aspects of the screening pathway have the potential to impact on mortality, overdiagnosis and diagnostic and treatment harms, if improved. A further strength is that the findings from the Delphi can be used to focus future research to provide evidence on aspects of the identified screening strategies that experts feel are important.

A limitation of the web-based format was the inability to have an in-depth response from respondents on the meaning and reasoning behind their answers as would be possible in a face-to-face interview setting, although in most cases the participants provided substantial and helpful comments. A further limitation is that the snowball approach may have led to researchers only recommending others they agree with, reinforcing any bias in the initial sample. There is also a concern that relevant potentially effective and cost-effective alternatives may have been missed or deliberately excluded by experts if they were not options that they themselves would support. We attempted to mitigate this by sending an evidence dossier to participants in advance to ensure they were aware of any relevant evidence and also by selecting participants with varying expertise and from a variety of institutions and countries to encourage diversity of opinions. We acknowledge, however, that there is still a potential risk of missing relevant aspects when using expert opinion rather than empirical evidence.

The choice to ask participants to indicate their preferred strategy in Round 1 did not allow us to estimate numerical uncertainty in their answers. However, participants were encouraged to comment on their answers, which is where any uncertainty was made clear. In the second round, participant’s uncertainty could be more clearly drawn out as they were requested to rank their agreement with the statements on a 9-point Likert scale. A further potential limitation is that the study aimed to achieve consensus on relevant screening strategies from participants from different countries with different health care systems. In general, the current standard of care in a particular health care setting may influence the set of comparators chosen, which could make incorporating an international perspective difficult. This is less of an issue in prostate cancer screening as current practice in the UK, Europe and the US is a generally consistent policy of shared decision making around whether or not a man undergoes PSA testing with a recommendation against formal screening. The limitations of an international perspective should be weighed against the benefits in any future applications, however.

Whilst the Delphi panel has identified screening strategies to assess in a cost-effectiveness analysis, there is limited or no trial evidence on several of the suggested strategies. This was highlighted by the participants throughout the process. When trial data are not available, as is often the case in screening due to the sample size required and associated expense, model-based economic evaluations are commonly used. Mathematical models can combine the best available sources of evidence on different aspects of screening to predict the costs and consequences of their use [32, 33]. However, it is important to take full account of the uncertainty in this evidence to prevent misguided conclusions. Any future cost-effectiveness model should consider

this uncertainty and enable adaption when new information becomes available.

5 Conclusion

The modified Delphi process identified six screening strategies that are relevant for comparing in a cost-effectiveness analysis against current practice (opportunistic screening). These strategies include no screening, age-based screening and different risk-stratified approaches, with all incorporating MRI. The work has demonstrated that the modified Delphi method could be used to identify relevant comparators for cost-effectiveness models when there is uncertainty due to rapidly evolving technologies. It may also be used to direct future research towards gathering evidence in the identified comparator pathways.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s40273-021-01009-6>.

Declarations

Funding This work was funded by Cancer Research UK and the UK Department of Health (C11043/A4286, C18281/A8145, C18281/A11326, and C18281/A15064, C18281/A24432). SS is funded by an NIHR fellowship (PDF-2017-10-008). HT and RMM were supported by the National Institute for Health Research (NIHR) Biomedical Research Centre at University Hospitals Bristol NHS Foundation Trust and the University of Bristol. The views expressed in this publication are those of the authors and not necessarily those of the NHS, the NIHR or the Department of Health and Social Care. The funding source was not involved in the design of the study, the collection, analysis and interpretation of data or the writing of the manuscript.

Conflict of interest RMM and ET report grants from Cancer Research UK during the conduct of the study. HT has received personal consulting fees from Novartis Pharma AG, Pfizer Inc, Roche Holding AG, and Janssen Pharmaceutica. The funding received by HT was not related to the present publication or in a similar clinical area. EK has received personal consulting fees from Novartis Pharma AG, Pfizer Inc, and Roche Holding AG. The funding received by EK was not related to the present publication or in a similar clinical area.

Author's contributions All authors were involved in the conception of the study. EK designed and administered the questionnaire and analysed the results with guidance from RM, ET, HT and SS. SS, HT, ET and RM contributed to the interpretation of the data and writing of the manuscript. EK wrote the first draft of the manuscript with SS, HT, ET and RM critically revising for important intellectual content. All authors read and approved the final manuscript.

Acknowledgements We acknowledge Peter Albertsen, Malcolm Mason, Sam Merriel, Anthony Zietman, David Neal, Nora Pashayan and Tyler Seibert for their advice and guidance during the initial development and piloting of the questionnaire. We also acknowledge Nora Pashayan, Henrik Gronberg, Anne Mackie, Peter Albertson, Divya Mallam, Sam Merriel, Hashim Ahmed, Fiona Walter, Michael Barry, Silvia Lombardo, Janet Rimmer, Allan Hackshaw, Tyler Seib-

ert, Anthony Zietman, Martin Eklund, Eveline A.M. Heijnsdijk, Anssi Auvinen, Caroline Moore, David Neal and Rosalind Eeles for participating in the consensus process.

Data availability Redacted datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Open Access This article is licensed under a Creative Commons Attribution-NonCommercial 4.0 International License, which permits any non-commercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

References

1. Sculpher M, Fenwick E, Claxton KJP. Assessing Quality in decision analytic cost-effectiveness models. *Pharmacoeconomics*. 2000;17(5):461–77.
2. Philips Z, et al. Good practice guidelines for decision-analytic modelling in health technology assessment. *Pharmacoeconomics*. 2006;24(4):355–71.
3. Yang Y, et al. Use of decision modelling in economic evaluations of diagnostic tests: an appraisal and review of health technology assessments in the UK. *Pharmacoeconomics Open*. 2019;3(3):281–91.
4. Tappenden P. Conceptual modelling for health economic model development. *HEDS Discussion Paper*. 2012;2012:12.
5. England PH. Prostate cancer risk management programme: benefits and risks of PSA testing. 2016.
6. Parker C, et al. Cancer of the prostate: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol*. 2015;26(5):v69–77.
7. Nam RK, et al. Prostate-specific antigen test for prostate cancer screening: American Society of Clinical Oncology provisional clinical opinion. *J Oncol Pract*. 2012;8(5):315.
8. Bibbins-Domingo K, Grossman DC, Curry SJJ. The US Preventive Services Task Force 2017 draft recommendation statement on screening for prostate cancer: an invitation to review and comment. *JAMA*. 2017;317(19):1949–50.
9. Smith RA, et al. Cancer screening in the United States, 2017: a review of current American Cancer Society guidelines and current issues in cancer screening. *CA Cancer J Clin*. 2017;67(2):100–21.
10. Wilt TJ, Harris RP, Qaseem A. Screening for cancer: advice for high-value care from the American College of Physicians. *Ann Internal Med*. 2020;162(10):718–25.
11. Ilic D, et al. Prostate cancer screening with prostate-specific antigen (PSA) test: a systematic review and meta-analysis. *BMJ*. 2018;362:k3519.
12. Seibert TM, et al. Polygenic hazard score to guide screening for aggressive prostate cancer: development and validation in large scale cohorts. *BMJ*. 2018;360:j5757.

13. Eklund M, Grönberg H, Nordström T. The STHLM3 prostate cancer diagnostic study: calibration, clarification, and comments. *Nature Rev Clin Oncol*. 2016;13(6):394.
14. Ahmed HU, et al. Diagnostic accuracy of multi-parametric MRI and TRUS biopsy in prostate cancer (PROMIS): a paired validating confirmatory study. *The Lancet*. 2017;389(10071):815–22.
15. Roberts M, et al. Conceptualizing a model: a report of the ISPOR-SMDM modeling good research practices task force–2. *Med Decision Making*. 2012;32(5):678–89.
16. Dalkey NC. The Delphi method: an experimental study of group opinion. 1969, RAND CORP SANTA MONICA CALIF. 1969.
17. Authority EFS. Guidance on expert knowledge elicitation in food and feed safety risk assessment. *EFSA J*. 2014;12(6):3734.
18. Landeta J. Current validity of the Delphi method in social sciences. *Technol Forecast Soc Chang*. 2006;73(5):467–82.
19. Avella JR. Delphi panels: research design, procedures, advantages, and challenges. *Int J Doctoral Stud*. 2016;11(1):305–21.
20. Patton MQ. *Qualitative research and evaluation methods*. Thousand Oaks: CA sage; 2002.
21. Kuper A, Lingard L, Levinson W. Critically appraising qualitative research. *BMJ*. 2008;337:a1035.
22. Harris PA, et al. Research electronic data capture (REDCap)—a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform*. 2009;42(2):377–81.
23. Akins RB, Tolson H, Cole BR. Stability of response characteristics of a Delphi panel: application of bootstrap data expansion. *BMC Med Res Methodol*. 2005;5:1–12.
24. Diamond IR, et al. Defining consensus: a systematic review recommends methodologic criteria for reporting of Delphi studies. *J Clin Epidemiol*. 2014;67(4):401–9.
25. Thorn JC, et al. Core items for a standardized resource use measure: expert Delphi consensus survey. *Value Health*. 2018;21(6):640–9.
26. Boulkedid R, et al. Using and reporting the Delphi method for selecting healthcare quality indicators: a systematic review. *PLoS ONE*. 2011;6(6):e20476.
27. Lindell MK, Brandt CJ, Whitney DJ. A revised index of interrater agreement for multi-item ratings of a single target. *Appl Psychol Meas*. 1999;23(2):127–35.
28. O'Neill TA. An overview of interrater agreement on likert scales for researchers and practitioners. *Front Psychol*. 2017;8:777–777.
29. Husbands S, et al. How qualitative methods can be used to inform model development. *Pharmacoeconomics*. 2017;35(6):607–12.
30. Iglesias CP, et al. Reporting guidelines for the use of expert judgement in model-based economic evaluations. *Pharmacoeconomics*. 2016;34(11):1161–72.
31. Sullivan W, Payne K. The appropriate elicitation of expert opinion in economic models. *Pharmacoeconomics*. 2011;29(6):455–9.
32. Irigorri N, Spackman E. Assessing the value of screening tools: reviewing the challenges and opportunities of cost-effectiveness analysis. *Public Health Rev*. 2018;39(1):17.
33. Sanghera S, et al. Cost-effectiveness of prostate cancer screening: a systematic review of decision-analytical models. *BMC Cancer*. 2018;18(1):84.